

Wolfgang Hoschek

Lawrence Berkeley National Laboratory
Distributed Systems Department / Secure Grid Technologies Group
1 Cyclotron Road, 50B-2239, Berkeley, CA 94720, USA
Phone: 415-515-6513 (mobile), 510-495-2460 (office), Fax: (510) 486 6363
Web: <http://dsd.lbl.gov/~hoschek/>
Email: whoschek@lbl.gov

Education

- Ph.D. Computer Science, Technical University of Vienna, Austria, Jan. 2003.
Dissertation: A Unified Peer-to-Peer Database Framework for XQueries over Dynamic Distributed Content and its Application for Scalable Service Discovery
Advisors: o. Univ.-Prof. Dipl.Ing. Mag. Dr. Gerti Kappel, ao. Univ.-Prof. Dr. Erich Schikuta, Dr. Bernd Panzer-Steindel
- M.S. Computer Science, University of Linz, Austria, 1997

Research Interests

- Structured and Unstructured XML Search, Fulltext Search, Metadata Repositories, Distributed Databases, Grid Computing, Peer-to-Peer Computing, Service Discovery, Systems Architecture and Design.

Research Experience

- **Postdoctoral Fellow at Berkeley Lab, 2/2003-present**
 - Designed and implemented an open-source Java toolkit for XML, XQuery, XQuery update, XPath, schema validation, fuzzy fulltext similarity search, Binary XML and related technologies. The tool is geared towards embedded use in high-throughput XML messaging middleware and main memory metadata repositories.
 - Specified, designed and implemented Peer-to-Peer Grid Service I/O architecture for access to dynamic data featuring ease-of-use, interoperability, scalability and performance. It is used to find, access and aggregate information and resources in a large distributed system composed of many autonomous data sources, dynamic and heterogeneous information, resources, participants and networks.
 - Extended prior work on Peer-to-Peer Grid Databases.
 - Designed and implemented a framework for a Staged Event Architecture, designed around non-blocking asynchronous communication facilities that are decoupled from the threading model chosen by any given application. Components for IP networking and in-memory communication are provided. It is used to easily build efficient and flexible low-level network clients and servers, and in particular as a basic communication substrate for Peer-to-Peer applications.
- **CERN Fellow, European Organization for Nuclear Research, 1/2001-12/2002**
 - Defined the overall European Data Grid architecture and standardized the design (joint work).
 - Defined the overall European Grid Data Management architecture and standardized the design of the European Grid Data Management Work Package (joint work).
 - Showed how to support expressive general-purpose service discovery queries over a view that integrates autonomous dynamic database nodes from a wide range of distributed system topologies spanning multiple administrative domains.
 - Designed and specified a database and query model as well as a generic and dynamic data model for service discovery.
 - Designed and specified a database (registry) that maintains information populated from a large variety of unreliable, frequently changing, autonomous and heterogeneous remote data sources, and that allows to express powerful queries over time-sensitive dynamic information.

- Designed and specified the Web Service Discovery Architecture (WSDA), which consists of a small set of orthogonal multi-purpose communication primitives (building blocks) for discovery. The individual primitives can be combined and plugged together by specific clients and services to yield a wide range of behaviors.
- Designed a Peer-to-Peer framework that is unified in the sense that it allows us to express specific applications for a wide range of data types, node topologies, query languages, query response modes, neighbor selection policies, pipelining characteristics, timeout and other scope options.
- Designed and specified a unified messaging, communication and network protocol model for Peer-to-Peer databases.
- Designed a general framework for creating replica location services that offer robust and scalable methods for identifying the storage systems on which specific replicas are located (joint work).
- Designed and implemented grid enabled relational database middleware (joint work).
- Designed Grid consistency service models that allow for replica update, synchronization and maintenance of different consistency levels (joint work).
- **CERN Doctoral Student, European Organization for Nuclear Research, 6/1998-2001**
 - Improved the performance of High-Energy Physics analysis through bitmap indices (joint work).
 - Designed and implemented uniform, versatile and efficient dense and sparse multi-dimensional arrays.
 - Designed and implemented a high performance analysis toolkit in Java.
 - Architected a Java binding for the CERN Espresso Object Database Management System.
 - Integrated and maintained an Open Source Distribution.
- **CERN Associate, European Organization for Nuclear Research, 1/1998-6/1998**
 - Prototyped efficient access methods for selections on Ntuples in Object Database Management Systems.
- **Summer Intern, Siemens-Nixdorf AG, Central Department of Research and Development (ZFE), Munich, Germany, 6/1995-10/1995**
 - Designed and prototyped a pretty printer for the Siemens Prolog product.

Teaching Experience

- Lector, University of Vienna, Institute for Computer Science, 2003.
- Assistant Lector, University of Linz, Austria, Department of Computer Science and Department of Software Engineering, 1992-1995

Publications

- A Unified Peer-to-Peer Database Framework for XQueries over Dynamic Distributed Content and its Application for Scalable Service Discovery. Wolfgang Hoschek, *PhD Thesis*, Technical University of Vienna, March 2002.
- Peer-to-Peer Grid Databases for Web Service Discovery. Wolfgang Hoschek, *Grid Computing: Making the Global Infrastructure a Reality*, Editors: Fran Berman, Geoffrey Fox and Tony Hey, March 2003, Wiley Press.
- Enabling Rich Service and Resource Discovery with a Database for Dynamic Distributed Content. Wolfgang Hoschek, *Journal of Parallel and Distributed Computing Practices (PDCP)*, special issue on Internet-based Computing, 2003.
- A Unified Peer-to-Peer Database Framework and its Application for Scalable Service Discovery. Wolfgang Hoschek, *Proc. of the 3rd Int'l. IEEE/ACM Workshop on Grid Computing (Grid'2002)*, Baltimore, USA, November 2002. Springer Verlag.
- The Web Service Discovery Architecture. Wolfgang Hoschek, *Proc. of the Int'l. IEEE/ACM Supercomputing Conference (SC 2002)*, Baltimore, USA, November 2002. IEEE Computer Society Press.
- Giggle: A Framework for Constructing Scalable Replica Location Services. Ann Chervenak, Ewa Deelman, Ian Foster, Leanne Guy, Wolfgang Hoschek, Adriana Iamnitchi, Carl Kesselman, Peter Kunszt, Matei Ripeanu, Bob Schwartzkopf, Heinz Stockinger, Kurt Stockinger and Brian Tierney, *Proc. of the Int'l. IEEE/ACM Supercomputing Conference (SC 2002)*, Baltimore, USA, November 2002. IEEE Computer Society Press.
- Web Service Discovery Processing Steps. Wolfgang Hoschek, *Proc. of the Int'l. WWW/Internet 2002 Conference*, Lisbon, Portugal, November 2002.
- A Comparison of Peer-to-Peer Query Response Modes. Wolfgang Hoschek, *Proc. of the Int'l. Conf. on Parallel and Distributed Computing and Systems (PDCS 2002)*, Cambridge, USA, November 2002.

- Dynamic Timeouts and Neighbor Selection Queries in Peer-to-Peer Networks. Wolfgang Hoschek, *Int'l. Conf. On Networks, Parallel and Distributed Processing and Applications (NPDPA 2002)*, Tsukuba, Japan, October 2002.
- Query Processing in Containers Hosting Virtual Peer-to-Peer Nodes. Wolfgang Hoschek, *Int'l. Conf. On Information Systems and Databases (ISDB 2002)*, Tokyo, Japan, September 2002.
- A Database for Dynamic Distributed Content and its Application for Service and Resource Discovery. Wolfgang Hoschek, *Int'l. IEEE Symposium on Parallel and Distributed Computing (ISPD 2002)*, Iasi, Romania, July 2002. **Best Paper Award.**
- Models for Replica Synchronisation and Consistency in a Data Grid. Dirk Duellmann, Wolfgang Hoschek, Javier Jean-Martinez, Asad Samar, Ben Segal, Heinz Stockinger and Kurt Stockinger, *10th IEEE Symposium on High Performance and Distributed Computing (HPDC-10)*, San Francisco, California, August 2001.
- Data Management in an International Data Grid Project. Wolfgang Hoschek, Javier Jaen-Martinez, Asad Samar, Heinz Stockinger and Kurt Stockinger, *1st IEEE/ACM Int'l. Workshop on Grid Computing (Grid'2000)*, Bangalore, India, December 2000. **Distinguished Paper Award.**
- Improving the Performance of High-Energy Physics Analysis through Bitmap Indices. Kurt Stockinger, Dirk Dullmann, Wolfgang Hoschek and Erich Schikuta, *11th Int'l. Conf. on Database and Expert Systems Applications*, London - Greenwich, UK, September 2000, Springer-Verlag.
- Project Spitfire - Towards Grid Web Service Databases. William Bell, Diana Bosio, Wolfgang Hoschek, Peter Kunszt, Gavin McCance, Mika Silander, *Global Grid Forum 5 Informational Document*, Edinburgh, Scotland, July 2002.
- A Unified Peer-to-Peer Database Protocol. Wolfgang Hoschek, *Data Grid TechReport DataGrid-02-TED-0407*, April 2002.
- A Data Model and Query Language for Service Discovery. Wolfgang Hoschek, *Data Grid TechReport DataGrid-02-TED-0409*, April 2002.
- Gigggle: A Framework for Constructing Scalable Replica Location Services. Ann Chervenak, Ewa Deelman, Ian Foster, Wolfgang Hoschek, Adriana Iamnitchi, Carl Kesselman, Peter Kunszt, Matei Ripeanu, Heinz Stockinger, Kurt Stockinger and Brian Tierney, *Global Grid Forum Working Draft*, Toronto, Canada, February 2002.
- Grid Enabled Relational Database Middleware. Wolfgang Hoschek and Gavin McCance, Glasgow University, *TechReport GLAS-PPE/2001-11*, November 2001 and *Global Grid Forum Informational Document*, Frascati, Italy, October 2001.
- The DataGrid Architecture. German Cancio, Steve M. Fisher, Tim Folkes, Francesco Giacomini, Wolfgang Hoschek, Dave Kelsey and Brian L. Tierney, *Data Grid TechReport DataGrid-ATF-01*, July 2001.

Professional Activities

- Member of the Grid Data Management Group of the European Data Grid Project (EDG).
- Member of two Global Grid Forum Working Groups (Open Grid Services Infrastructure and Grid Information Services).
- Served on the European Data Grid Architecture Taskforce and as the European Data Grid - Globus Contact Person for DataGrid and Java Interfaces.
- Referee for *Computer Communications* and *Journal of Concurrency and Computation: Practice and Experience*.
- Member of IEEE and ACM, SIGMOD.

Software Projects

- Nux - Open-source Java toolkit for XML, XQuery, XQuery update, XPath, schema validation, fuzzy fulltext similarity search, Binary XML and related technologies (architect and lead developer).
- Firefish – Java Peer-to-Peer Grid service infrastructure for access to dynamic data featuring ease-of-use, interoperability, scalability and performance (architect and lead developer).
- Staged Event Architecture (SEA) - Non-blocking asynchronous communication facilities that are decoupled from the threading model chosen by any given application (architect and lead developer).
- European Data Grid (EDG) - Next generation computing infrastructure providing computation and analysis of shared large-scale databases, from hundreds of TeraBytes to PetaBytes, across widely distributed scientific communities (project member).
- Spitfire - A set of grid enabled database middleware services (architect and lead developer).
- Colt - Open Source Libraries for High Performance Scientific and Technical Computing in Java (architect, lead developer and maintainer).

- CERN Java Infrastructure - maintainer of an extensive Java infrastructure for some hundred products and thousands of users.